CALIF 4.0
THE INTERACTIVE SHINY WEB APPLICATION FOR CALIBRATION OF SURVEY WEIGHTS

UNECE Workshop on Measuring Poverty in Pandemic Times

26.03.2021

Boris Frankovič

Statistical Office of the Slovak Republic

Contents

- calibration essentials
- optimal strategy
- Calif 4.0
- results for Slovak EU-SILC

Calibration approach

- population U, sample S
- every unit in S with design weight d_k as the inverse probability of selection
- nonresponse occurs and accounts for non-sampling error – weights need to be adjusted for it
- without adjustment they cannot reproduce neither known population totals X_j nor estimated characterstics precisely

Calibration approach

- seeking for new weights w_k for each unit in S that confirm X_j and differ minimally from initial weights
- then these new weights are used for estimation – they treat some level of nonresponse bias
- found by optimization methods
- difference between initial and calibration weights computed by distance functions with parameter $r_k = \frac{w_k}{d_k}$
- bounds for the differences can be set (some feasible interval)

Calibration approach

- proposed by Deville and Särndal
- significant enhancement of precision of estimates
- study variables should correlate with auxiliary calibration variables
- when response rates are unequally distributed and cause bias
- brings consistence among surveys

Optimal strategy

- common rule has been to find such solution that bounds are as strict as possible
- this stemmed from perception of a tiny difference between initial and calibration weights when they are kept within some tense interval
- however, limited space forces the initial weights towards bounds and as a result the average weight difference is quite high

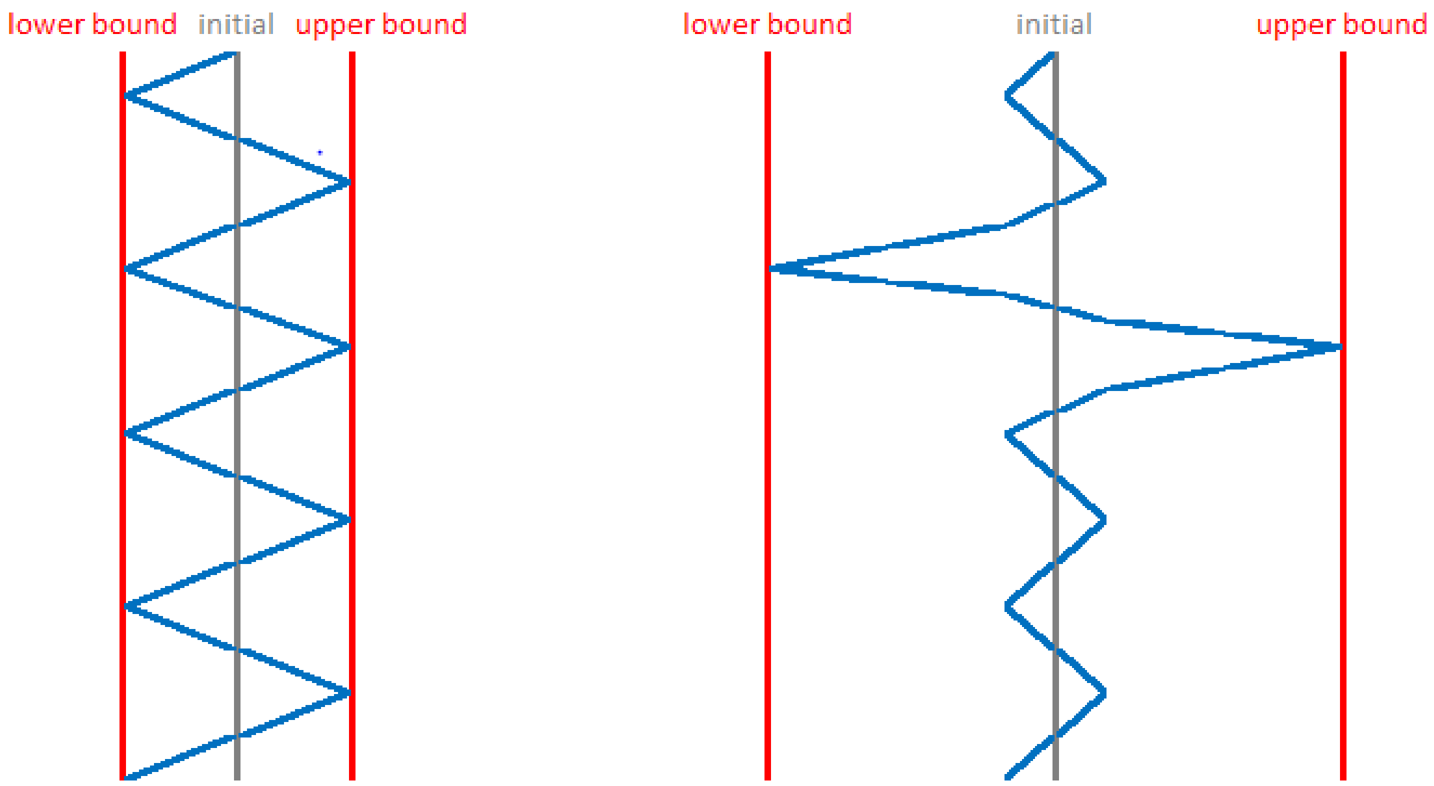
Optimal strategy

- less strict constraints form some area to move, where just a small number of weights is pushed off to the bounds while others remain close to their origins

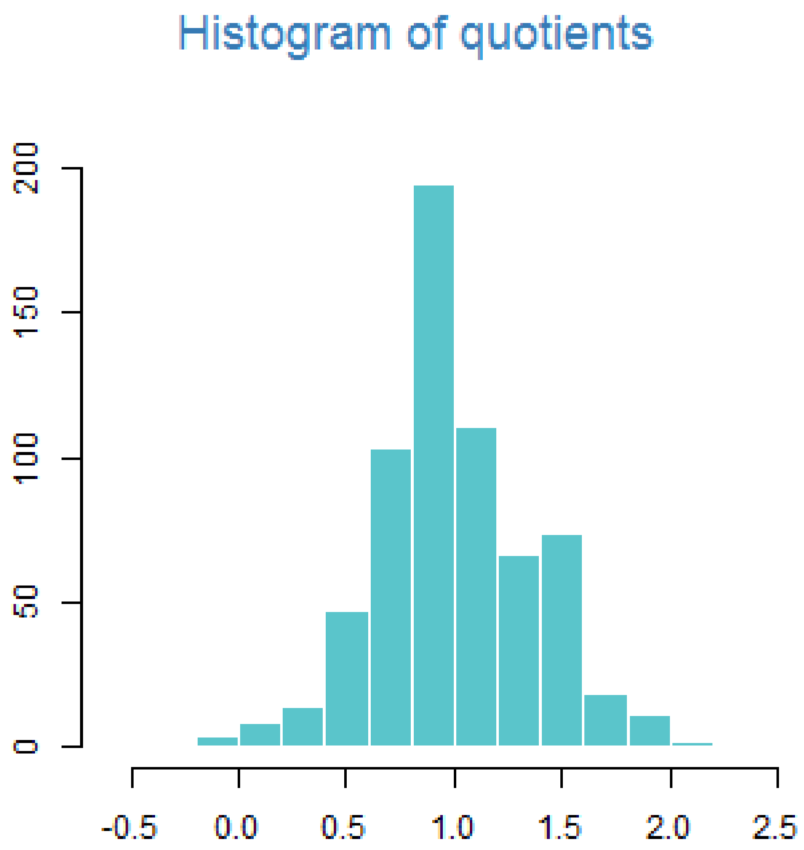
- the average difference between initial and calibration weights is thus minimal for unbounded methods

- to avoid extreme weights, some bounded method should be used, but a feasible trade-off between the width of the interval and the distortion applied has to be found

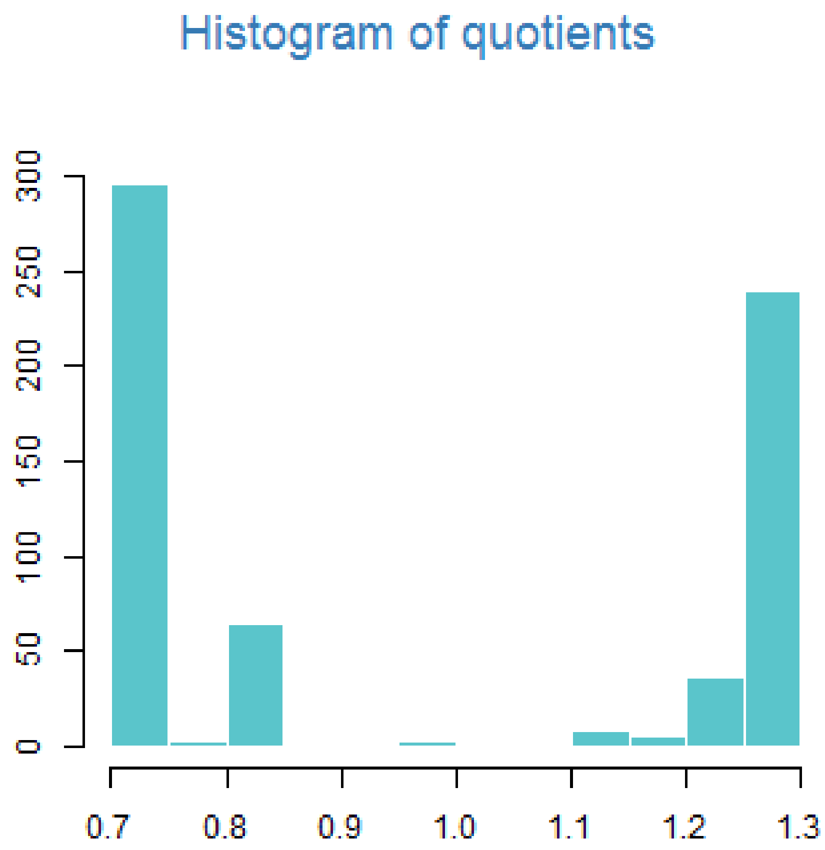
Optimal strategy



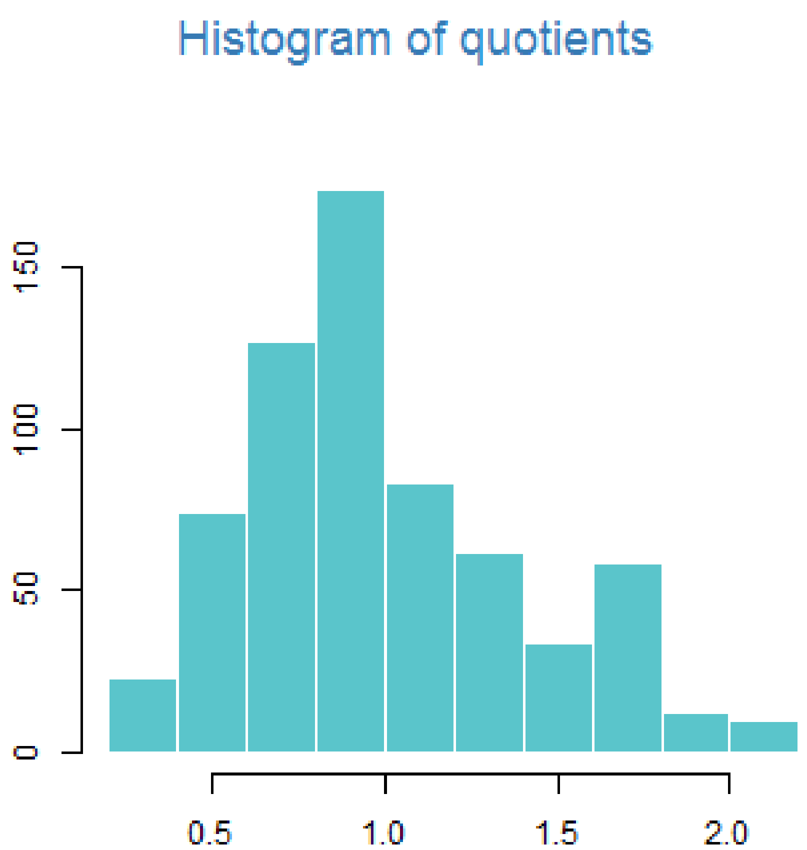
Weight quotients – no bounds



Weight quotients – strict bounds



Weight quotients – loose bounds – feasible trade-off



Optimal strategy - conditions

- calibration constraints should be reproduced
- calibration weights differ not very much from initial weights
- they should lay in rather tense interval
- the average weight difference should be low compared to the linear method

Calif 4.0

- shiny web application – in R
- GPL v3.0 licence
- modern design
- easy to work with, very fast, maintaining time-proven techniques
- web browser act as a display device
- all computations run locally in R – no data are sent to the Internet
- https://github.com/SO-SR/Calif
- https://slovak.statistics.sk/wps/portal/ext/products/software.tools

Main features of Calif

- web-designed GUI
- convenient operability
- two-stage calibration for integrated weights
- interactive data loading
- stratification
- detailed and user-defined specification of calibration variables
- powerful and fast solvers

Main features of Calif

- commonly used distance functions
- approximate solutions possible
- bookmarking and interactive saving
- graphical and worthful outputs
- and many others

Installation

- install R – free
- install required packages

```
install.packages(c('shiny', 'haven', 'sampling', 'nleqslv'))
```

- go to github or SO SR's webpage and run respective command, either

```
shiny::runGitHub('Calif', 'SO-SR', destdir = getwd(),
launch.browser = TRUE)
```

- or

```
shiny::runUrl('https://slovak.statistics.sk/wps/wcm/connect/
7014bfd4-54a2-4080-929f-eb949bf25e39/calif.zip?MOD=
AJPERES&CVID=m7Xjumj&CVID=m7Xjumj', filetype = '.zip',
destdir = getwd(), launch.browser = TRUE)
```

Overview tab

Calif 4.0 | Overview | Data | Calibration

Welcome to Calif!

Calif is the Shiny web application used for calibration of weights in statistical surveys. It is an open-source software and you are free to use it, also without any knowledge of R programming language.

Detailed information on how to use Calif can be found in the Manual

Getting started

To get started working with Calif, you first need to prepare the data and the table of totals (which are the known population marginals). You can learn it in the Manual, in fact it is very easy and intuitive.

Once they are ready, you need to load them into Calif in the next tab After that, you select the calibration variables, either categorical or numerical, initial weights for calibration, possible stratification and calibration methods.

Optimal strategy

Then you can calibrate the weights. In Calif, after each calibration step, you are provided with useful statistics that can help you to decide which method and parameters are most suitable for performing the best possible calibration.

The optimal strategy is to find the new weights that reproduce the population totals and are as close as possible to the initial weights such that the lower and upper bounds form a narrow interval (optimally close to 1 from both sides)

and the average difference between the initial and calibration weights is low in comparison to the linear solution. The feasibility of the average difference is demonstrated by a pie chart on the **Calibration** tab.

If you set too tense bounds, even if the solution is found, the histogram of quotients on the **Calibration** tab will look unnaturally, the average difference will be high and such calibration will not be appropriate.

Working directory

Your actual working directory is set to `D:/Users/frank2/Desktop` . If you want to change it, you can do it by clicking on this button

Bookmarking

You can bookmark your work at any time by clicking on the **Bookmark** button on the **Calibration** tab. The files and settings are saved into your working directory and you obtain a link that can be used to reproduce your work when Calif is run again.

Output

You can save your work at any time by clicking on the **Save** button on the **Calibration** tab. The output is an identical copy of your data with column *CalibrationWeight* added. This column represents your latest calibration attempt, i.e.

if you tried several calibrations with different settings, the last one is recorded. If you haven't performed any calibration yet, *CalibrationWeight* column contains the initial weights.

If you carry out stratified calibration, *CalibrationWeight* column contains the new weights for strata that have been already calibrated and the initial weights for strata that have remained untouched.

Although the latest state of play is still remembered throughout all strata, you are recommended to regularly save or bookmark your work in order not to lose it if some unexpected error occurs.

Data tab

Period Comma

Specify calibration variables

☑ Stratification

Weights: Weight

Stratification variable: REGION

Numerical variables

s1a6
s2a1
s2a2
s2a3
s2a4
s2a5
s2a6
MEMBERS
Weight

Categorical variables

s1a6
s2a1
s2a2
s2a3
s2a4
s2a5
s2a6
MEMBERS
Weight

Other settings

Indicators monitoring (optional)

id_hd
REGION
s1a1
s1a2
s1a3

☐ Total (%)
☐ Total (value)
☐ Mean (%)
☐ Mean (value)

Missings

Missings for categorical, separated with commas

☐ Eliminate rows with missings

Maximum nr of iterations

20000

Tolerance

0,00001

Proceed

Dataset with 6 rows and 18 columns

Show 5 entries

Search:

STRA	MEMBERS_1	MEMBERS_2	MEMBERS_3	MEMBERS_4	MEMBERS_5	s1a1	s1a2	s1a3	s1a4	s1a5	s1a6	s2a1	s2a2	s2a3	s2a4	s2
1	64841	58791	48099	48019	15489	45907	34665	99262	39020	38788	29654	43365	33689	101594	45010	472
3	50982	48547	40169	47186	26081	43774	42256	94469	43697	37300	29888	41784	40310	89073	43526	407
4	70510	59378	48030	57141	26692	52052	47017	109497	50025	43107	32739	48754	44584	104615	50507	492
5	55249	53313	42573	49341	37621	60506	52184	109673	48127	39322	28641	57549	50221	103446	47445	436
6	64808	59340	47747	53156	21460	53872	45809	103172	47450	39862	29400	51399	44327	99249	48260	452
STRA	MEMBERS_1	MEMBERS_2	MEMBERS_3	MEMBERS_4	MEMBERS_5	s1a1	s1a2	s1a3	s1a4	s1a5	s1a6	s2a1	s2a2	s2a3	s2a4	s2a

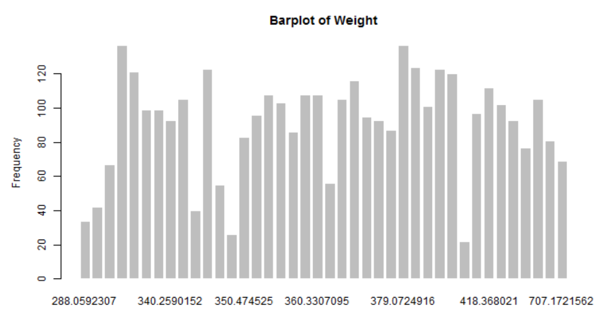
Showing 1 to 5 of 6 entries

Previous 1 2 Next

Explore variables

Weight

☑ Explore as categorical

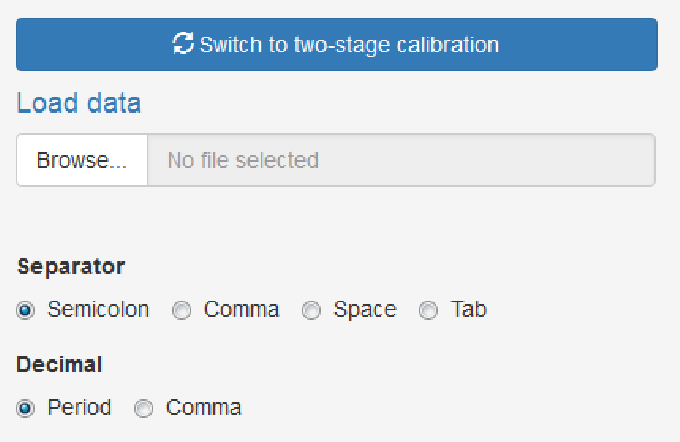


Value	Frequency
288.0592307	34
300.0748022	42
317.0452195	67
327.6444278	137
334.3032222	121
339.8685987	99
339.9882421	99
340.2590152	93
340.7154621	105
345.0188148	40
346.7000684	123

STATISTICAL OFFICE OF THE SLOVAK REPUBLIC

Data tab

- import of data and table of totals
- .csv, .txt and .sas7bdat formats supported
- data interactively displayed
- switch between single-stage and two-stage calibration



Data tab

- data structure can be almost free – it can contain any variables, but at least calibration variables and initial weights
- required structure for the table of totals is described in the Manual
- after the calibration, data with the same structure are returned, with one column added

Calibration tab

Calif 4.0 | Overview | Data | Calibration

Choose strata

1
3
4
5
6
2
7
8

Show with initial weights

Method & Solver

Method	Solver
○ Linear	◉ calib
○ Raking ratio	○ nleqslv
○ Logit	
◉ Linear bounded	

Lower bound	Upper bound
0,3	2

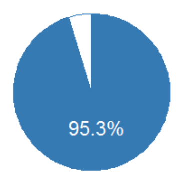
CALIBRATE

Bookmark | Save

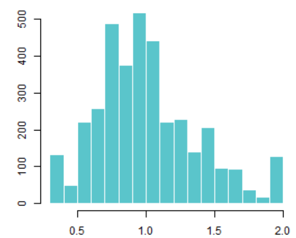
Results

Initial weights interval	288.059	707.172
Calibration weights interval	86.418	1260.294
Lower bound obtained	0.300	
Upper bound obtained	2.000	
Average weight quotient	1.008	
Average difference	105.647	
Minimum realistic lower bound	0.004	

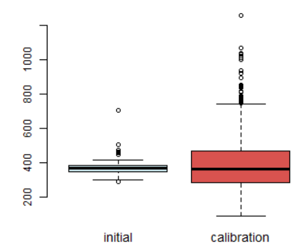
Average difference feasibility



Histogram of quotients



Boxplots of weights



Totals obtained

☐ Show obtained totals as values

Stratum	MEMBERS_1	MEMBERS_2	MEMBERS_3	MEMBERS_4	MEMBERS_5	s1a1	s1a2	s1a3	s1a4	s1a5	s1a6
1	100.00%	100.00%	100.00%	100.00%	100.00%	97.88%	100.00%	100.00%	100.00%	100.00%	100.00%
3	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
4	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
6	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Weights & Quotients

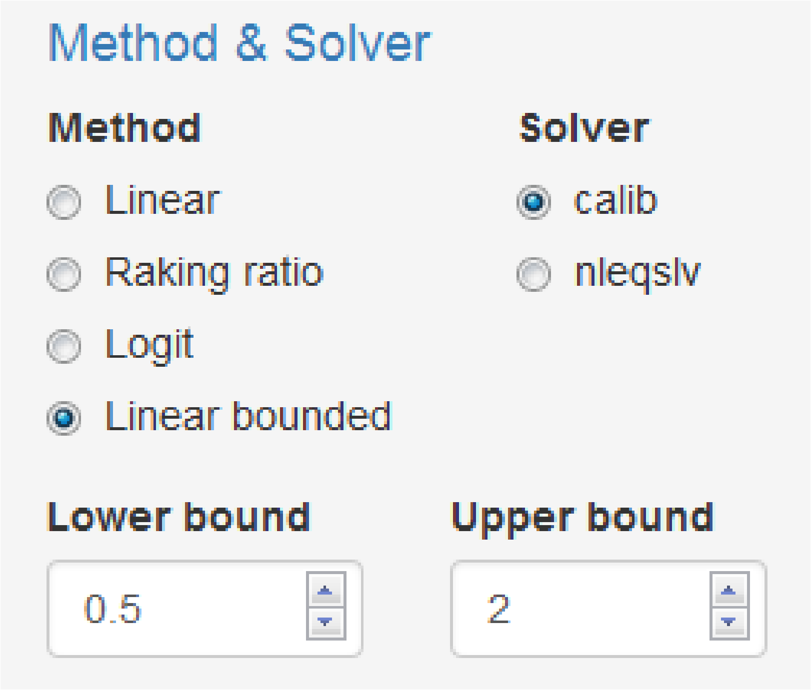
Show 20 entries

Row	Initial	Calibration	Quotients
3266	288.05923	86.41777	0.3

Calibration tab

- selection of strata, proper method and solver
- show with initial weights – H-T estimates of calibrated totals using initial weights as compared to these totals – difference between actual reality and desired target – very useful tool to examine distribution of nonresponse in the survey – values far away from 100% are subject to non-sampling error due to nonresponse
- *calib* + *linear bounded* ideal for social surveys (such as EU-SILC)

Calibration tab



Calibration tab

- summary statistics – obtained totals, former and latter weight intervals, attained weight quotients, average difference $(\mathrm{AD} = \frac{1}{n}\sum_{k=1}^{n}|w_k - d_k|)$

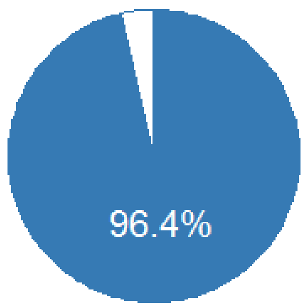
Results

Initial weights interval	286.604	337.132
Calibration weights interval	85.981	761.426
Lower bound obtained	0.300	
Upper bound obtained	2.657	
Average weight quotient	1.000	
Average difference	98.330	
Minimum realistic lower bound	0.004	

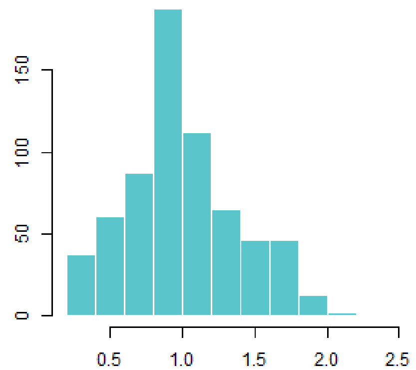
Calibration tab

- useful figures

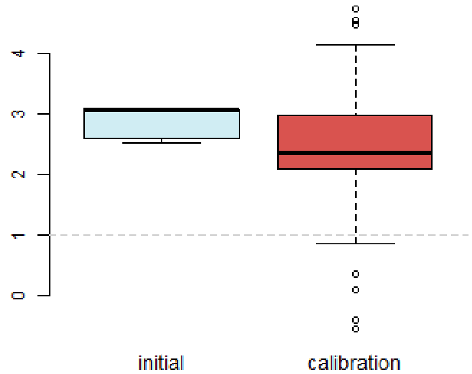
Average difference feasibility



Histogram of quotients



Boxplots of weights



EU-SILC 2016 results

- each new panel calibrated separately on the whole population
- new 2016 panel – 1510 households
- calibration criteria separately in each NUTS3 level
  - 8 sex-age grous
  - employees
  - unemployed
  - pensioners
  - 2 household size categories

EU-SILC 2016 results

- altogether 104 totals to calibrate on
- resulting totals 100% reproduced with further settings

NUTS3	Solver	Method	Lower bound	Upper bound
1	calib	linear bounded	0,3	3
2	calib	linear bounded	0,3	2,7
3	calib	linear bounded	0,3	2,2
4	calib	linear bounded	0,3	2,2
5	calib	linear bounded	0,4	2,2
6	calib	linear bounded	0,2	2,8
7	calib	linear bounded	0,3	2,5
8	calib	linear bounded	0,3	2,2

EU-SILC 2016 results

- the whole cross-sectional file calibrated once more
- 2016 cross-sectional file – 5738 households
- calibration criteria separately in each NUTS3 level
    - 12 sex-age grous
    - employees
    - self-employed
    - unemployed
    - pensioners
    - 2 household size categories

STATISTICAL OFFICE OF THE SLOVAK REPUBLIC

EU-SILC 2016 results

- altogether 144 totals to calibrate on
- resulting totals 100% reproduced with further settings

NUTS3	Solver	Method	Lower bound	Upper bound
1	calib	linear bounded	0,5	1,5
2	calib	linear bounded	0,5	1,4
3	calib	linear bounded	0,5	1,3
4	calib	linear bounded	0,5	1,5
5	calib	linear bounded	0,5	1,3
6	calib	linear bounded	0,5	1,3
7	calib	linear bounded	0,5	1,4
8	calib	linear bounded	0,5	1,3

Thank you for your attention